

# An Optimal Transportation based Univariate Neuroimaging Index

Liang Mi<sup>1</sup>   Wen Zhang<sup>1</sup>   Junwei Zhang<sup>2</sup>   Yonghui Fan<sup>1</sup>   Dhruvan Goradia<sup>3</sup>  
 Kewei Chen<sup>3</sup>   Eric M. Reiman<sup>3</sup>   Xianfeng Gu<sup>2</sup>   Yalin Wang<sup>1</sup>  
<sup>1</sup>Arizona State University   <sup>2</sup>Stony Brook University   <sup>3</sup>Banner Alzheimer's Institute  
<http://gsl.lab.asu.edu>.

## Abstract

*The alterations of brain structures and functions have been considered closely correlated to the change of cognitive performance due to neurodegenerative diseases such as Alzheimer's disease. In this paper, we introduce a variational framework to compute the optimal transformation (OT) in 3D space and propose a univariate neuroimaging index based on OT to measure such alterations. We compute the OT from each image to a template and measure the Wasserstein distance between them. By comparing the distances from all the images to the common template, we obtain a concise and informative index for each image. Our framework makes use of the Newton's method, which reduces the computational cost and enables itself to be applicable to large-scale datasets. The proposed work is a generic approach and thus may be applicable to various volumetric brain images, including structural magnetic resonance (sMR) and fluorodeoxyglucose positron emission tomography (FDG-PET) images. In the classification between Alzheimer's disease patients and healthy controls, our method achieves an accuracy of 82.30% on the Alzheimers Disease Neuroimaging Initiative (ADNI) baseline sMRI dataset and outperforms several other indices. On FDG-PET dataset, we boost the accuracy to 88.37% by leveraging pairwise Wasserstein distances. In a longitudinal study, we obtain a 5% significance with  $p$ -value =  $1.13 \times 10^{-5}$  in a  $t$ -test on FDG-PET. The results demonstrate a great potential of the proposed index for neuroimage analysis and the precision medicine research.*

## 1. Introduction

Neuroimaging studies the measurement of brain structures and functions and their changes during development, aging, learning, disease, and evolution [27]. Modern techniques analyze noninvasive neuroimaging data typically obtained from structural Magnetic Resonance Imaging (sMRI) and fluorodeoxyglucose positron emission tomography (FDG-PET) by using mathematical and statistical

methods. A ubiquitous approach for neuroimaging research is population based analysis, e.g. voxel based morphometry (VBM) [46] [5] [45] [7] and tensor based morphometry (TBM) [16] [12] [44]. They are being widely used to identify differences in the local composition of brain tissue while discounting large-scale differences in gross anatomy and position [26]. VBM/TBM studies the brain abnormality by non-linearly registering the brain images and comparing the local differences among brain tissues and their functions voxel by voxel in order to obtain the statistical significance between groups and to discover the general trend of disease burden and progression. The past decade has witnessed a growing interest in developing new techniques to automatically quantify the abnormality of brain structure and functions, especially for the study of neurodegenerative diseases such as Alzheimer's disease (AD) [10] [40].

However, prior works (e.g. [38]) have shown that the statistically significant voxels obtained from the group difference studies do not necessarily carry strong statistical power for computer-aided diagnosis (CAD). In clinical settings, doctors prefer single indices as biomarkers because single indices best quantify the brain structures and functions for individuals rather than in the group level which has less clinical effectiveness. A single neuroimaging index provides a practical reference for evaluating disease progression and for effective treatments. Yet the automated analysis of brain diseases (e.g. AD) based on single indices faces many challenges largely because these single indices can hardly represent the complicated shape or metabolic activity of the entire human cerebral cortex. Most researchers focus on developing indices from region of interest (ROI) study (e.g. [10]). However, the ROI-based techniques require prior knowledge and precise segmentation which is a major limitation that could result in inconsistency across different studies [42]. In randomized clinical trials (RCT), the regulatory agencies prefer single valued outcome measure which does not require statistical correction for multiple regional comparisons. Moreover, functional imaging techniques such as FDG-PET often suffer from low resolution and noises, which prevents traditional methods from

accurate analysis. Therefore, it is highly advantageous for clinical diagnosis and prognosis that we develop a unified, robust method that works on different modalities to quantify the shape of the cerebral cortex with a single index.

We have observed that after segmentation and normalization, 3D brain images can be treated as 3D probability distributions in the Euclidean domain, where (1) the intensity of each voxel is non-negative, (2) the total intensity is one, and (3) the distance between two voxels is the Euclidean distance. Therefore, the problem of comparing brain images can be converted to the problem of comparing the corresponding distributions. The Wasserstein distance, which is a numerical value, measures the similarity between two probability distributions in a given metric space. It is obtained by solving the underlying optimal transportation problem which is to measure the minimum cost for moving the total mass of one distribution to another. Thus, optimal transportation (OT) is often referred to as optimal mass transport (OMT) and so is the Wasserstein distance as the earth mover’s distance. Recently, the Wasserstein distance has proved itself to be a robust shape descriptor for 3D surfaces. Rabin *et al* computed the Wasserstein distance on 3D surfaces to perform shape retrieval [31]. Solomon *et al* proposed the convolutional Wasserstein distance and explored many of its applications including shape interpolation and skeleton layout [36]. In [37] and [33], the Wasserstein distance were applied to brain surfaces to find the correlation between IQ and brain shapes, and Alzheimer’s disease and brain shapes, respectively. Lévy [25] generalized a surface optimal transportation algorithm to 3D and solved it with quasi-Newton method. ur Rehman *et al* [39] solved OT on the sMRI for 3D registration, which reveals the feasibility to compute the Wasserstein distance on 3D brain volumes.

Despite of the convenience of using the Wasserstein distance to measure distributions, a limiting factor is the computational complexity of solving the optimal transportation problem. This is especially true when dealing with high dimensional distributions, and is the reason that ur Rehman *et al* [39] adopted GPU to improve efficiency. In practice, some methods seek to approximate the optimal transportation in order to avoid heavy computation, e.g. [36] [15].

In this paper, we introduce a numerically efficient algorithm to compute the optimal transportation with Newton’s method. Based on the algorithm, we propose a framework for indexing brain images. First, we normalize the images and re-discretize them in order to obtain their tetrahedral representations. Then, we compute the optimal transportation between the tetrahedral meshes and obtain the 3D Euclidean Wasserstein distance between them. Because of its properties as a metric and its sensitivity to the difference between two measures, the Wasserstein distance servers as a univariate neuroimaging index to distinguish between individual images and between images in the group level. We

apply our framework on two modalities of brain images – sMRI and FDG-PET – and verify the Wasserstein distance as a univariate index in distinguishing Alzheimer’s disease patients from cognitively normal people. Our results show that the Wasserstein distance is a highly stable feature and provides a relatively high classification accuracy of 82.30% on sMRI and 88.37% on FDG-PET with boosting.

The contribution of this research is threefold. First and foremost, we explore the potential of the Wasserstein distance as a univariate neuroimaging index in the analysis of both sMR and FDG-PET images. Second, we provide an alternative to score the changes in brain tissue and their functions from their appearance on brain images. We model the brain tissues and functions with their intrinsic geometric structure and obtain a shape distance between each pair. To our knowledge, this is the first time that a geometric approach has been proposed to quantify the whole brain change on FDG-PET. Third, we propose a framework based on the variational optimal transportation to compute the 3D Euclidean Wasserstein distances of brain images. The variational framework is efficient, which makes it advantageous for working with large-scale and high dimensional datasets.

## 2. Preliminaries

We first introduce the problem of optimal transportation in the context of brain image analysis. We also describe volumetric harmonic mapping in 2.2, which provides a convex canonical space for computing optimal transportation .

### 2.1. Optimal Transportation

Suppose  $M(\mathbf{m}, \mu)$  and  $N(\mathbf{n}, \nu)$  are two metric spaces with probability measures. In other words,  $1 = \int_M \mu(m)dm = \int_N \nu(n)dn$ . We call a mapping  $T : M(\mathbf{m}, \mu) \rightarrow N(\mathbf{n}, \nu)$  a measure preserving one if the measure of any subset  $B$  of  $N$  is equal to the measure of the origin of subset  $B$  in  $M$ , which means  $\mu(T^{-1}(B)) = \nu(B)$ ,  $\forall B \subset N$ . Given a transportation cost function  $c : M \times N \rightarrow \mathbb{R}_0^+$ , the problem of optimal transportation (OT) is to find the measure preserving mapping  $T_{opt} : M \rightarrow N$  that minimizes the total cost,

$$T_{opt}(M, N) = \arg \min_T \int_M c(m, T(m))d\mu(m). \quad (1)$$

For simplicity,  $T_{opt}$  is replaced by  $T$  in the rest of the paper.

In the context of brain mapping under discrete settings, given two brain images  $M(\mathbf{m}, \mu)$  and  $N(\mathbf{n}, \nu)$ , the problem becomes to find the OT,  $T : M \rightarrow N$ , such that the cost is the minimum. And we call the minimum cost the Wasserstein distance between  $M$  and  $N$ , which is defined as:

$$w(M, N) := \inf \sum_{m \in M} c(m, T(m))\mu(m). \quad (2)$$

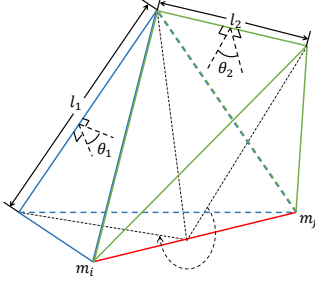


Figure 1: Edge  $(m_i, m_j)$  (red) is shared by two tetrahedra (blue and green). Each tetrahedron has an edge and dihedral angle against edge  $(m_i, m_j)$ . The summation of the length  $l_t$  times dihedral angle  $\theta_t$  forms the harmonic string energy.

The cost  $c(\cdot)$  is defined as the quadratic Euclidean distance between a voxel and its map,  $c(m, T(m)) = \|m - T(m)\|^2$ . The Wasserstein distance is a metric that quantifies the similarity of two domains which in our case are two brain images. The details of the optimal transportation problem and the properties of the Wasserstein distance can be found in [19] [23] [41] and [8].

## 2.2. Volumetric Harmonic Mapping

A tetrahedral mesh  $M$  can be regarded as a simplicial complex  $K$  under embedding  $f : |K| \rightarrow \mathbb{R}^3$ . Suppose  $m_i$  and  $m_j$  are two adjacent vertices in the mesh, and  $f(m_i)$  and  $f(m_j)$  are their coordinates in  $\mathbb{R}^3$ . Then, our goal is to find the  $f'$  that minimizes a string energy,

$$E(f) \stackrel{\text{def.}}{=} \langle f, f \rangle = \sum_{(m_i, m_j) \in K} k(m_i, m_j) \|f(m_i) - f(m_j)\|^2. \quad (3)$$

$k(m_i, m_j)$  is called the string constant. Different definitions of  $k$  determine different types of string energy. Suppose an edge  $(m_i, m_j)$  is shared by  $n$  tetrahedra. Then we define the harmonic string constant as

$$k(m_i, m_j) \stackrel{\text{def.}}{=} \frac{1}{12} \sum_{t=1}^n l_t \cot(\theta_t), \quad (4)$$

where  $l_t$  is the length of the edge to which edge  $(m_i, m_j)$  is against and  $\theta_t$  denotes the dihedral angle on that edge. Figure 1 depicts the geometric relations. The red line represents edge  $(m_i, m_j)$  shared by several tetrahedra, two of which (one blue, one green) are shown in the figure.  $\theta_t$ ,  $t = 1, 2$  is the dihedral angle in Equation (4) and so is the edge length  $l_t$ . Plugging Equation (4) into Equation (3) gives us the harmonic energy. We apply gradient descent to get  $f'$ . [43] provides details on volumetric harmonic mapping. Briefly, we first compute a brain surface spherical

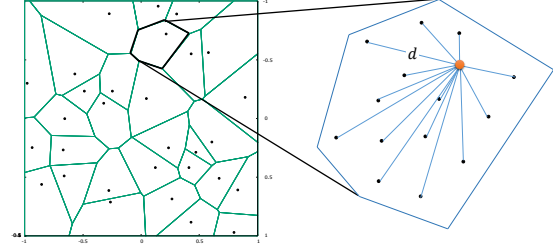


Figure 2: Voronoi diagram on a 2D plane. The plane is partitioned into different cells, each with a particle. Every measure point on the plane has a smaller distance to the particle of its own cell than to any other particle.

conformal mapping. Using the sphere as a fixed boundary, we minimize the volumetric harmonic energy (Equation 3) and produce the volumetric harmonic mapping. After volumetric harmonic mapping, an arbitrary tetrahedral mesh will be harmonically mapped into a unit sphere. This is the initial condition to perform the discrete optimal transportation which requires a convex metric space.

## 3. Discrete Variational Optimal Transportation

Having introduced the optimal transportation problem in 2.1, given a discrete point set with a Dirac measure  $(p, \nu) = \{(p_i, \nu_i) | p_i \in \mathbb{R}^n, \sum_i \nu_i = 1\}$ ,  $i = 1, 2, \dots, l$ , our goal is to find a discrete optimal transportation from a metric space  $M(m, \mu) \subset \mathbb{R}^n$  to the discrete point set  $(p, \nu)$  which partitions the metric space into polygons, forming a power Voronoi diagram. Figure 2 illustrates a Voronoi diagram on a 2D Euclidean plane. Black dots represent the measure  $(m, \mu)$  while the orange dots indicate the point set  $(p, \nu)$ . We introduce a height vector  $\mathbf{h} = (h_1, h_2, \dots, h_l)$  and a piecewise linear convex function:

$$u_{\mathbf{h}}(m) = \max\{\langle m, p_i \rangle - h_i\}, i = 1, 2, \dots, l. \quad (5)$$

Thus, each Voronoi cell can be represented as:

$$V_i \stackrel{\text{def.}}{=} \{m \in M \mid d^2(p_i, m) - \tilde{h}_i \leq d^2(p_j, m) - \tilde{h}_j\}, \forall j \neq i, \quad (6)$$

which is called a *power Voronoi cell* since we compute the power distance  $d^2(p_i, m) - \tilde{h}_i$  where

$$\tilde{h}_i = 2h_i - |p_i|^2.$$

The gradient map  $\nabla u_{\mathbf{h}} : V_i(\mathbf{h}) \rightarrow p_i$  “transports” each  $V_i(\mathbf{h})$  to a specific point  $p_i$ . The total mass of  $V_i(\mathbf{h})$  is denoted as:  $w_i(\mathbf{h}) = \mu(V_i(\mathbf{h}))$ . For a given metric space  $M(m, \mu)$  and any given Dirac measure  $(p, \nu)$ , according

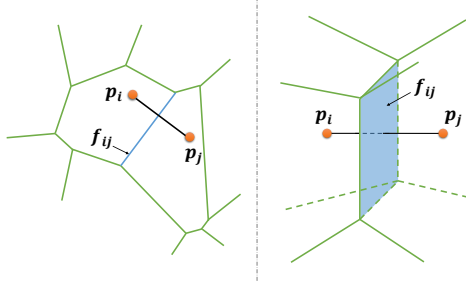


Figure 3: The intersection (blue) of Voronoi cells in 2D (left) and 3D (right) cases. The hessian matrix is derived from the geometric relation between adjacent Voronoi cells.

to [20], there exists a unique height vector  $\mathbf{h}$  such that the gradient map  $\mu(\mathbf{h})$  is measure preserving, i.e.  $w_i(\mathbf{h}) = v_i$  and meanwhile minimized the quadratic transportation cost.

Now, we define an energy function,

$$E(\mathbf{h}) \stackrel{\text{def}}{=} \int^{\mathbf{h}(\mathbf{h})} \sum_{i=1}^l \mathbf{h} w_i d\mathbf{h}_i - \sum_{i=1}^l \nu_i h_i. \quad (7)$$

Equation (7) is a convex function, which makes its optimization tractable via Newton’s method. We follow the variational principle and derive gradient and Hessian matrix. The gradient of the energy is

$$\nabla E(\mathbf{h}) = (w_1(\mathbf{h}) - \nu_1, \dots, w_l(\mathbf{h}) - \nu_l)^T. \quad (8)$$

The Hessian matrix can be computed from the power

---

**Algorithm 1:** Computing discrete variational optimal transportation

---

**Data:** A convex set

$(M, \mu) = \{(m_1, \mu_1), (m_2, \mu_2), \dots, (m_k, \mu_k)\}$ ,  
a set of particles with a Dirac measure

$(\mathbf{p}, \nu) = \{(p_1, \nu_1), (p_2, \nu_2), \dots, (p_l, \nu_l)\}$ ,  
a threshold  $\epsilon$ , and a step  $\lambda$ .

**Result:** Discrete optimal transportation map

$\varphi : (M, \mu) \rightarrow (\mathbf{p}, \nu)$ , represented as  $(V, \mathbf{h})$ .

```

1 begin
2    $\mathbf{h} \leftarrow (0, 0, \dots, 0)$ .
3   repeat
4     Compute  $V$  with current  $(\mathbf{p}, \mathbf{h})$ .
5     Compute  $\mathbf{w}(\mathbf{h}) = \{w_i = \sum_{m \in V_i} \mu(m)\}$ .
6     Compute  $\nabla E(\mathbf{h})$  using Equation (8).
7     Compute  $H$  using Equation (9).
8      $\mathbf{h} \leftarrow \mathbf{h} - \lambda H^{-1} \nabla E(\mathbf{h})$ .
9   until  $|\nabla E(\mathbf{h})| < \epsilon$ 
10  return  $\varphi$ .
11 end
```

---

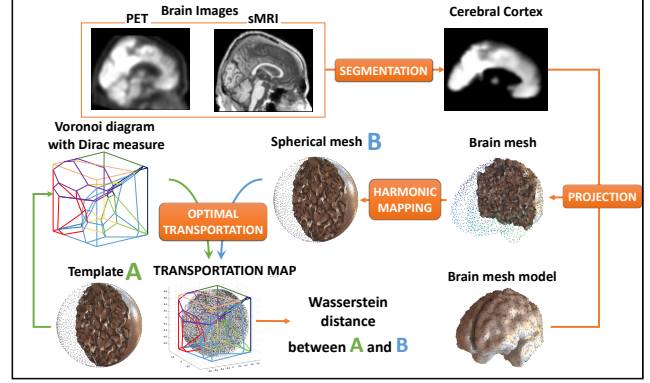


Figure 4: The pipeline to compute the Wasserstein distance index of an sMR or FDG-PET image. It consists segmentation, projection, harmonic mapping, and optimal transportation. Eventually, the output of the framework is the Wasserstein distance between mesh A (green) and mesh B (blue).

Voronoi diagram. Specifically, suppose two cells  $V_i(\mathbf{h})$  and  $V_j(\mathbf{h})$  intersect at a hyperplane  $f_{ij} = V_i(\mathbf{h}) \cap V_j(\mathbf{h})$ , as shown in Figure 3. The Hessian matrix of  $E(\mathbf{h})$  is given by

$$H \stackrel{\text{def}}{=} \left( \frac{\partial^2 E(\mathbf{h})}{\partial h_i \partial h_j} \right) = \left( \frac{\text{Area}(f_{ij})}{|p_j - p_i|} \right). \quad (9)$$

By Newton’s method, at each step, we solve a linear system,

$$H \delta \mathbf{h} = \nabla E(\mathbf{h}), \quad (10)$$

and update the height vector  $\mathbf{h} \leftarrow \mathbf{h} - \lambda \delta \mathbf{h}$  until  $\delta \mathbf{h}$  is below a pre-defined threshold. The step  $\lambda$  is an empirical value. The initial value of  $\mathbf{h}$  is set to  $\mathbf{h} \leftarrow (0, 0, \dots, 0)^T$ . Letting  $\varphi$  denote the optimal transportation map  $\varphi : M(\mathbf{m}, \mu) \rightarrow (\mathbf{p}, \mathbf{h})$ , we show the complete algorithm for computing the optimal transportation map  $\varphi$  in Algorithm 1.

## 4. OT based Neuroimaging Index

Figure 4 shows the overview of our framework. It shows the data flow from the brain images to its Wasserstein distance index. The flow before “Projection” is considered pre-processing which is introduced in 4.1. In 4.2, we talk about how to transform the 3D digitalized image into a tetrahedral representation. After that, we describe our approach to obtain a neuroimaging index for each brain volume via discrete optimal transportation (OT) in 4.3.

### 4.1. Data Pre-processing

In this study, we focus on two modalities of brain images – structural magnetic resonance image (sMRI) and fluorodeoxyglucose positron emission tomography (FDG-PET) image. Due to different mechanism of these two modalities, we approach to pre-processing them in different ways and in here we describe them separately.



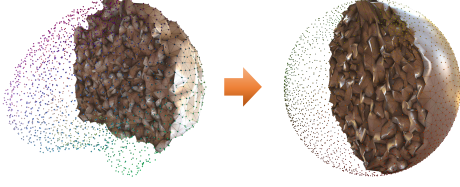


Figure 5: Volumetric harmonic mapping from a brain mesh to a unit spherical mesh. The feature (e.g. intensity) at each vertex remains the same after the mapping.

**sMRI** The sMRI scans we use for each subject are T1-weighted which have been acquired at 1.5T. First, we linearly align all the images together using FLIRT [35] which is a linear image registration tool included in FM-RIB Software Library (FSL) [1]. After that, we use FreeSurfer [17][2] to process all the images to obtain the segmented brain tissues (e.g. cerebral cortex) and statistics on the brain structure (e.g. average cortical thickness).

**FDG-PET** The FDG-PET scans used in this research are processed as follows. First, we use the software toolkit Statistical Parametric Mapping (SPM) [30][3] to linearly align the images into a common space. Second, we borrow a brain mask from SPM, manually exclude the brain stem and only keep cerebral cortex (because cerebral cortex is the object of this study), and then segment all the images with this cerebral mask. Third, we conduct spatial smoothing with a Gaussian kernel of the full width at half maximum (FWHM) equal to (8, 8, 8) in three directions ( $x, y, z$ ).

## 4.2. Re-discretization

The purpose of re-discretization is to replace the images with regularized tetrahedral meshes so that it is easier to perform shape transformation. In this study, we adopt the segmented brain volumes from the pre-processing phase and use FreeSurfer for sMRI and SPM on FDG-PET respectively to extract their surfaces. Then, we use TetGen [34] to generate the tetrahedral meshes from the surfaces.

Given a brain image and its corresponding tetrahedral mesh aligned together, we project the intensities of the images onto the meshes. Suppose, after the alignment, a voxel of the image is surrounded by several vertices of the mesh. First, for each voxel  $x$  of the image, we use ANN (Approximate Nearest Neighbor) library [29] to find its four nearest vertices,  $m = \{m_i\}, i = 1, \dots, 4$ , on the mesh. The number of the nearest points we count for projection is adjustable to the difference in the resolutions between the image and the mesh. After that, the intensity  $I(x)$  is dispensed from  $x$  to  $m_i$ . The value  $\mu(m_i)$  that  $m_i$  receives is determined by the

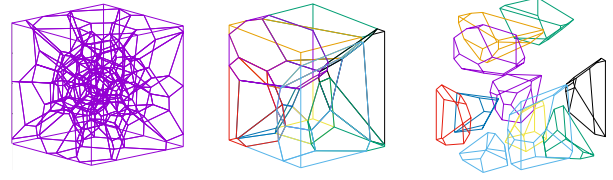


Figure 6: 3D Voronoi diagrams. As the same as in 2D scenario, the points in each 3D weighted Voronoi cell has the smallest weighted distance to its own particles than to any other particle.

distance between that  $m_i$  and  $x$ . In particular, we have:

$$\mu(m_i) = a_i I(x), \quad (11)$$

$$a_i = \frac{1/d_i}{\sum_{i=1}^4 1/d_i}, \quad (12)$$

where  $d_i = \sqrt{\|x - v_i\|^2}$  denotes the Euclidean distance between  $x$  and  $v_i$ . The closer vertex receives a larger value of intensity. This ensures the smoothness of the data.

Another way of determining  $a_i$  is through inverse barycentric method, but inverse barycentric could result in negative values if the “center” is outside the convex hull formed by neighbors, especially in the regions close to the cerebral cortex surface. A negative vertex feature does not have a physical meaning. Hence, we do not apply inverse barycentric method, although it is mathematically correct.

## 4.3. Discrete Optimal Transportation

**Volumetric Harmonic Mapping** In order to provide the convexity that the optimal transportation requires, we employ volumetric harmonic mapping which maps an arbitrary 3D tetrahedral mesh to a unit sphere. Figure 5 illustrates a brain mesh before and after the harmonic mapping. Colors indicate the intensities that have been projected from the original images as the result of re-discretization.

**Discrete Optimal Transportation** After building the common canonical space by volumetric harmonic mapping, we apply Algorithm 1 to compute the optimal transportation between desired brain meshes. Figure 6 (left) shows the resulting Voronoi diagram with 61 cells which is what we use in the experiments. After the optimal transportation, the size of each Voronoi cell has been adjusted so that each cell encloses the vertices with measures whose total value approximates the Dirac measure of the center of the cell. To show the shape of each Voronoi cell, we also draw the central 11 cells of the diagram (middle and right) with different colors. We implement our algorithm in C/C++ and we borrow Voropp [32] to compute the 3D Voronoi diagram. We solve the linear system in Equation (10) using the square conjugate gradient solver provided by Eigen library [4].

Table 1: Demographic information of subjects in this study

Modality	Study	Cohort	#	Age*	MMSE*
sMRI	Cross.	AD	75	76.3	23.1
		CN	175	76.1	29.0
FDG-PET	Cross.	AD	23	76.4	18.3
		CN	20	78.4	28.8
	Long.	AD	30×2	77.8	18.2
		CN	30×2	79.1	28.4

Key: AD Alzheimer’s disease; CN cognitively normal; Cross. Cross-sectional; FDG-PET fluorodeoxyglucose positron emission tomography; Long. Longitudinal; MMSE Mini-Mental State Examination; sMRI structural magnetic resonance imaging; \* mean values.

**3D Euclidean Wasserstein Distance** Recall Equation (2) defining the minimum cost brought by the optimal transportation. We call this minimal cost 3D Euclidean Wasserstein distance if the quadratic distance function is defined in the 3D Euclidean space where the brain meshes are embedded in. We describe our approach for computing the Wasserstein distance index in the rest of the paragraph. First, we build a template from a brain mesh. Specifically, we initialize an unweighted Voronoi diagram with point set,  $P(\mathbf{p}, \nu)$  where  $\nu_i = 0, \forall i$ . The points are manually designed to be uniformly distributed in the canonical space which is the harmonic map of the brain mesh as shown in Figure 4. After that, we calculate total mass of each Voronoi cell, which is the summation of the intensities of all the vertices in this particular cell, and assign it to the measure of the corresponding point. Because the total mass of the entire mesh is one, the newly generated point set is actually a Dirac measure. The point set will serve as the template for computing the optimal transportation between the brain mesh model and each other brain mesh  $M(\mathbf{m}, \mu)$  using Algorithm 1. The Wasserstein distance is then computed by

$$w(M, P) = \sum_{m_i=1}^k d^2(m_i, p_j) \mu_i, \quad (13)$$

where  $p_j = \varphi(m_i)$  and  $d^2(m_i, p_j) = \|m_i - p_j\|^2$ .

Because the Wasserstein distance measures the similarity between the measurement of different brain meshes, by comparing the distances from all the meshes to a common template, we will obtain a one dimensional distance space  $W$ , called the Wasserstein space. The origin is the Wasserstein distance between the template and itself and the coordinates serve as single indices representing the shape of all the brain volumes regarding to that template.

Table 2: Timing of computing the Wasserstein distance between two tetrahedral meshes

	# of vertices	18,614
	# of tetrahedra	111,466
Newton’s	# of iteration	176
	Time	54.85s
Gradient descent	# of iteration	1,441
	Time	397.21s

## 5. Applications to FDG-PET and sMR Images for Alzheimer’s Disease Analysis

In this section, we use four separate experiments to evaluate the performance of our indexing framework on sMR and FDG-PET images for the analysis of Alzheimer’s disease. The test data come from Alzheimers Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu) [22] [21]. We randomly selected 164 FDG-PET scans from ADNI-2 and 250 sMRI scans from ADNI-1 databases. The detailed cohort information is described in Table 1. In the rest of the paper, we refer to FDG-PET as PET.

### 5.1. Timing and Numerics

All the tests were run on a 3.40 GHz Intel (R) Core (TM) i7-3770 CPU with 8.00 GB RAM. To achieve a threshold of  $3.2 \times 10^{-3}$  (convergence) on a tetrahedral mesh with 18,614 vertices, the average running time of our method for computing the Wasserstein distance between two tetrahedral meshes over 10 trials was 54.85 seconds with 176 iterations. Step  $\lambda$  was 0.3. Table 2 summaries the timing and numerics. Our approach via Newton’s method outperforms the traditional method based on gradient descent [39]. Both were tested without GPU or CPU multi-threaded acceleration – the performance can be further improved. The results provide us a practical foundation to study the Wasserstein distance as a univariate neuroimaging index.

### 5.2. Cross-sectional Study of AD on sMRI

To explore the practicality of our framework on sMR images as well as its robustness over large brain image datasets, we apply it to the ADNI-1 baseline dataset with 75 AD and 175 CN subjects, and compare the Wasserstein distance index with four frequently used single indices – hippocampal volume, average entorhinal cortex thickness, average cortical thickness and total grey matter volume [13] [6] [14] – in terms of their classification accuracies. To make them comparable, we use the minimum number of “base classifiers” – two base classifiers. One is based on an AD template and the other is based on a CN template.

First, we randomly select 10 AD and 10 CN images

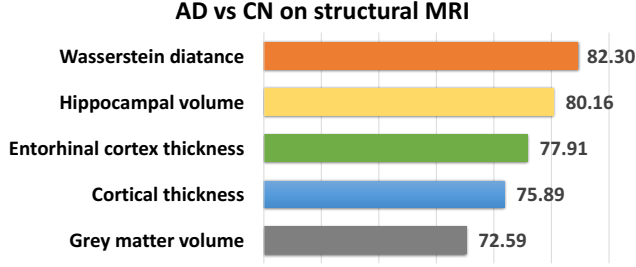


Figure 7: Results from the classification between AD and CN on ADNI-1 baseline. Among all the indicators, the Wasserstein distance achieves the highest accuracy.

and respectively average them into two templates, each for a group. Then, we compute the Wasserstein distances of each baseline image to these 2 templates. Thus each image will obtain 2 values. We regard these 2 values as a 2-dimensional feature and run linear discriminant analysis (LDA) on them. Finally, we use the statistics computed from FreeSurfer [17] to evaluate the performance of our proposed method. We conduct the classification with 10-fold cross-validation. Each time, we adopt 90% of the whole dataset with their true labels (AD or CN) to train a classification model and then apply this model to the remaining 10% to obtain the result. Figure 7 summarizes the performance of different indices. Among all the indices, the Wasserstein distance index achieves the highest accuracy of 82.30%. Since the Wasserstein distance index works on a large ADNI baseline dataset, we expect it as a potential index to be useful in assisting clinical diagnosis.

### 5.3. Cross-sectional Study of AD on FDG-PET

Prior research has shown strong evidence that PET is a valid prognostic marker of progression of Alzheimer’s disease [28]. Geometric analysis of PET, however, has been limited due to low resolution and lack of anatomical information [24][9]. Our framework regards each image as a probability measure and does not require the positional relation between adjacent voxels, which makes it applicable to PET images. We compute the Wasserstein distance between each two subjects. According to our previous discussion, in the group level, the subjects in the same cohort (AD or CN) may have similar brain functions compared to the brain functions of subjects in different cohorts. We test our hypothesis through classification between AD and CN.

Our approach is inspired by AdaBoost [18]. To be detailed, each time we regard an image  $I_i$  as a template and compute the Wasserstein distance  $w(I_i, I_j)$  from itself to all others  $I_j, j \neq i$ , and repeat the process over all the images.

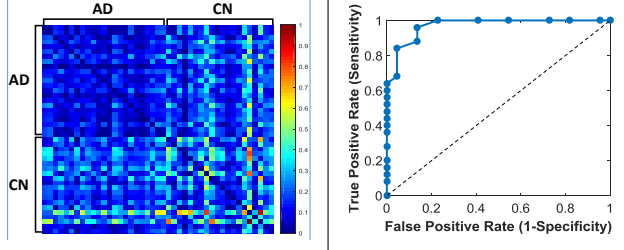


Figure 8: Distance matrix obtained from computing the Wasserstein distance between each two FDG-PET images.

Figure 9: ROC curve for classification between AD and CN on FDG-PET images.

Therefore, each time we get a base classifier

$$h_i(\mathbf{w}_j^i) = \begin{cases} \text{sign}(I_i), & \text{if } \mathbf{w}_j^i < w_{\text{threshold}} \\ -\text{sign}(I_i), & \text{if } \mathbf{w}_j^i \geq w_{\text{threshold}} \end{cases} \quad (14)$$

where  $\text{sign}(I_i) \in \{1, -1\}$  indicates the class of  $I_i$ , which is the ground truth, and  $\mathbf{w}_j^i$  is the Wasserstein distance between image  $I_j$  and template  $I_i$ . The final classifier is

$$H(\mathbf{w}_j) = \text{sign} \left( \sum_{i=1}^N \alpha_i h_i(\mathbf{w}_j^i) \right), i \neq j, \quad (15)$$

where  $\mathbf{w}_j$  represents the distance vector by computing the Wasserstein distances between image  $I_j$  and all the templates including itself (but we exclude itself as the template when we compute the output of the classifier). Since we know the label (AD or CN) of the template, by setting a proper threshold,  $w_{\text{threshold}}$ , we can obtain the labels of all the other images. Due to the small number of the data available to us, we consider that each base classifier equally contributes to the final classifier, i.e.  $\alpha_i = \frac{1}{N}, \forall i$ .

Figure 8 shows the Wasserstein distance matrix,  $\mathbf{w}$ , a collection of all the distance vectors  $\mathbf{w}_j$ , each representing a column (or a row). The distances have been normalized to the range [0 1] for the purpose of visualization. By leveraging our approach to classify 23 AD patients and 20 CN subjects, we obtained an accuracy of 88.37% with sensitivity of 91.30% and specificity of 85.00%. The area under curve (AUC) is 97% as shown in Figure 9.

An interesting point we can find from Figure 8 is that although the images labeled “AD” can be clustered by their Wasserstein distances very well, images labeled as “CN” are relatively not clustered together but rather distributed over the space. One explanation could be, that when people have Alzheimer’s disease often associated with the decline in metabolic rate they form a similar pattern in terms of the Wasserstein distance. It is well known that AD is highly correlated to certain common regions across different subjects. We treat each brain volume as a distribution. Therefore, we hypothesize that the 3D distributions will become

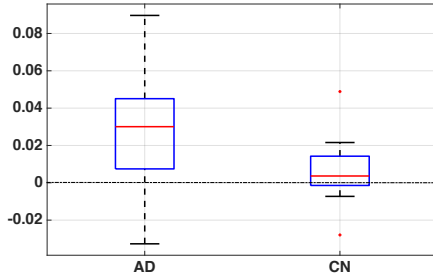


Figure 10: Box plot indicating the distribution of the Wasserstein distances of AD and CN.

closer in terms of their pair-wise Wasserstein distances because the loss of the brain tissue (e.g. empty or almost-empty voxels) during the progression of AD will gain more weights in the distribution. As a result, the Wasserstein distances between AD images become smaller but others do not have significant change.

#### 5.4. Longitudinal Study of AD on FDG-PET: A Statistical Significance Test

Clinical findings have shown strong association between metabolic reduction and the cognitive decline among patients with AD [11]. Therefore, in the time domain, the Wasserstein distance measure between the PET images and the template should become smaller (assuming we have an AD template) since the distribution of the images becomes more similar to the template. In addition, the Wasserstein distances of the PET images labeled AD should decrease at a higher rate than the images of the CN subjects do.

We verify our hypothesis with a longitudinal experiment involving 30 AD patients and 30 CN subjects each having 2 scans spanning 2 years. First, for each PET image, we calculate its Wasserstein distance to every other PET image labeled “AD”, and then average the results. This averaged Wasserstein distance will serve as a one-dimensional feature for the rest of the analysis. Second, for each subject with two PET scans, we record the change in the Wasserstein distance of the PET scan from initial visit to 2-year visit. Therefore, we hypothesize that for AD subjects, the change should be larger than that of CN subjects because AD patients have fast metabolic reduction in cerebral cortex as opposed to normal people whose cerebral metabolic activities do not show significant change. Figure 10 reveals the box plot of the longitudinal experiment. The one-side t-test yields a 5% significance with  $p\text{-value} = 1.13 \times 10^{-5}$ , which very well accords with our hypothesis. Our statistical results show that the proposed univariate index is a promising neuroimaging biomarker for tracking AD progression and measuring responses to interventions.

## 6. Conclusions

Optimal transportation has long been studied in the areas of computer graphics and vision where most research has been focused on 2D images and 3D surfaces. The complexity of the underlying optimization has been a challenge to researchers especially for high dimensional data, e.g. 3D volumes. Our proposed algorithm for computing variational optimal transportation provides a solution to process high dimensional data within reasonable time. We proposed a novel framework based on variational optimal transportation to compute the Wasserstein distance between brain images, and demonstrated the potential of the Wasserstein distance as a brain imaging index in characterizing neurodegenerative diseases. Our framework for indexing brain images provides a new perspective to analyze brain images. We solve the optimal transportation problem on different modalities of brain images for characterizing brain structures and metabolic patterns, and obtain promising results that verify our hypothesis: people with Alzheimer’s disease tend to have similar brain structures and metabolic activity patterns in terms of their Wasserstein distances.

There are two open issues with our framework. One is to determine the optimal number of Voronoi cells used in optimal transportation. Using more cells will likely increase the accuracy of the mapping but will also boost the computational cost. Thus, it would be ideal to discover an optimal number of cells that balances efficiency and effectiveness. The other issue comes from the selection of templates. In 5.2, we adopt two templates instead of one because finding an ideal template for OT is extremely challenging and is still an open problem. If a non-typical volume (with smaller Wasserstein distances to most AD volumes but belongs to CN) is selected as the template, the predictions can be far from the truth and even the opposite. We choose to use two templates to reduce the bias. More empirical knowledge has to be discovered for building a high-quality template that can work alone for classification tasks.

Our work provides a tool for solving the optimal transportation problem efficiently in high dimensional spaces and explores its applications to characterizing Alzheimer’s disease on sMR and FDG-PET images. Future work could include solving the OT problem for even higher dimensional data, e.g. tensor images. Additionally, the use of the proposed index to other modalities, such as functional and diffusion MRI, may further improve pre-symptomatic diagnosis and treatment of neurodegenerative diseases.

**Acknowledgements** This work was partially supported by the National Institutes of Health (R21AG043760, RF1AG051710 and U54EB020403) and National Science Foundation (DMS-1413417 and IIS-1421165). The first author would also like to thank Dr. Jingrui He from Arizona State University, USA and Ms. Jing Wang from Northwestern University, USA for their suggestions on classification.



## References

- [1] <http://fsl.fmrib.ox.ac.uk/>. 5
- [2] <http://freemurfer.net/>. 5
- [3] <http://www.fil.ion.ucl.ac.uk/spm/>. 5
- [4] <http://eigen.tuxfamily.org/>. 5
- [5] F. Abell, M. Krams, R. P. John Ashburner, K. Friston, R. Frackowiak, F. Happ, C. Frith, and U. Frith. The neuroanatomy of autism: a voxelbased whole brain analysis of structural scans. *Neuroreport*, 10(8):1647–1651, 1999. 1
- [6] S. K. Andersen, C. E. Jakobsen, C. H. Pedersen, A. M. Rasmussen, M. Plocharski, and L. R. stergaard. Classification of alzheimers disease from mri using sulcal morphology. In *Scandinavian Conference on Image Analysis*, pages 103–113, 2015. 6
- [7] J. Ashburner and K. J. Friston. Voxel-based morphometry—the methods. *Neuroimage*, 11(6 Pt 1):805–821, Jun 2000. 1
- [8] Y. Brenier. Polar factorization and monotone rearrangement of vectorvalued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991. 3
- [9] O. Camara, G. Delso, O. Colliot, A. Moreno-Ingelmo, and I. Bloch. Explicit incorporation of prior anatomical information into a nonrigid registration of thoracic and abdominal ct and 18-fdg whole-body emission pet images. *IEEE transactions on medical imaging*, 26(2):164–178, 2007. 7
- [10] K. Chen, N. Ayutyanont, J. B. Langbaum, A. S. Fleisher, C. Reschke, W. Lee, X. Liu, D. Bandy, G. E. Alexander, P. M. Thompson, L. Shaw, J. Q. Trojanowski, C. R. Jack, S. M. Landau, N. L. Foster, D. J. Harvey, M. W. Weiner, R. A. Koeppe, W. J. Jagust, and E. M. Reiman. Characterizing Alzheimer’s disease using a hypometabolic convergence index. *Neuroimage*, 56(1):52–60, May 2011. 1
- [11] K. Chen, J. B. Langbaum, A. S. Fleisher, N. Ayutyanont, C. Reschke, W. Lee, X. Liu, D. Bandy, G. E. Alexander, P. M. Thompson, et al. Twelve-month metabolic declines in probable alzheimer’s disease and amnesic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the alzheimer’s disease neuroimaging initiative. *Neuroimage*, 51(2):654–664, 2010. 8
- [12] M. Chung, K. Dalton, L. Shen, A. Evans, and R. Davidson. Weighted Fourier representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging*, 26:566C–581, 2007. 1
- [13] M. J. Clarkson, M. J. Cardoso, G. R. Ridgway, M. Modat, J. D. R. Kelvin K. Leung, N. C. Fox, and S. Ourselin. A comparison of voxel and surface based cortical thickness estimation methods. *Neuroimage*, 57.3:856–865, 2011. 6
- [14] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M.-O. Habert, M. Chupin, H. Benali, O. Colliot, A. D. N. Initiative, et al. Automatic classification of patients with alzheimer’s disease from structural mri: a comparison of ten methods using the adni database. *neuroimage*, 56(2):766–781, 2011. 6
- [15] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013. 2
- [16] C. Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *Comput. Vis. Image Understanding*, 66:207–222, 1997. 1
- [17] B. Fischl. Freesurfer. *Neuroimage*, 62.2:774–781, 2012. 5, 7
- [18] Y. Freund and R. E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory*, pages 23–37. Springer, 1995. 7
- [19] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002. 3
- [20] X. Gu, F. Luo, J. Sun, and S.-T. Yau. Variational principles for minkowski type problems, discrete optimal transport, and discrete monge-ampere equations. *arXiv preprint arXiv:1302.5472*, 2013. 4
- [21] C. R. Jack, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner. The Alzheimer’s Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging*, 27(4):685–691, Apr 2008. 6
- [22] W. J. Jagust, D. Bandy, K. Chen, N. L. Foster, S. M. Landau, C. A. Mathis, J. C. Price, E. M. Reiman, D. Skovronsky, R. A. Koeppe, et al. The alzheimer’s disease neuroimaging initiative positron emission tomography core. *Alzheimer’s & Dementia*, 6(3):221–229, 2010. 6
- [23] L. V. Kantorovich. On a problem of Monge. *Journal of Mathematical Sciences*, 133(4):1383–1383, 2006. 3
- [24] Z. Keidar, N. Haim, L. Guralnik, M. Wollner, R. Bar-Shalom, A. Ben-Nun, and O. Israel. Pet/ct using 18f-fdg in suspected lung cancer recurrence: diagnostic value and impact on patient management. *Journal of Nuclear Medicine*, 45(10):1640–1646, 2004. 7
- [25] Lvy, Bruno. A numerical algorithm for l2 semi-discrete optimal transport in 3d. *ESAIM: M2AN*, 49(6):1693–1715, 2015. 2
- [26] A. Mechelli, C. J. Price, K. J. Friston, and J. Ashburner. Voxel-based morphometry of the human brain: methods and applications. *Current medical imaging reviews*, 1(2):105–113, 2005. 1
- [27] D. Mietchen and C. Gaser. Computational morphometry for detecting changes in brain structure due to development, aging, learning, disease and evolution. *Frontiers in neuroinformatics*, 3(25), 2009. 1
- [28] L. Mosconi. Brain glucose metabolism in the early and specific diagnosis of Alzheimer’s disease. FDG-PET studies in MCI and AD. *Eur. J. Nucl. Med. Mol. Imaging*, 32(4):486–510, Apr 2005. 7
- [29] D. Mount and S. Arya. Ann: library for approximate nearest neighbour searching. 1998. 5

- [30] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Academic press, 2011. 5
- [31] J. Rabin, G. Peyr, and L. D. Cohen. Geodesic shape retrieval via optimal mass transport. In *European Conference on Computer Vision. Springer Berlin Heidelberg*, pages 771–784, 2010. 2
- [32] C. Rycroft. A three-dimensional voronoi cell library in c++. *Lawrence Berkeley National Laboratory*, 2009. 5
- [33] J. Shi, W. Zhang, and Y. Wang. Shape analysis with hyperbolic wasserstein distance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5051–5061, 2016. 2
- [34] H. Si. Tetgen, a delaunay-based quality tetrahedral mesh generator. *ACM Transactions on Mathematical Software (TOMS)*, 41(2):11, 2015. 5
- [35] S. Smith, M. Jenkinson, M. Woolrich, C. Beckmann, T. Behrens, H. Johansen-Berg, P. Bannister, M. De Luca, I. Drobnjak, D. Flitney, and R. Niazy. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage*, 23:S208–S219, 2004. 5
- [36] J. Solomon, F. D. Goes, G. Peyr, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (TOG)*, 24(4):66, 2015. 2
- [37] Z. Su, W. Zeng, Y. Wang, Z.-L. Lu, and X. Gu. Shape classification using wasserstein distance for brain morphometry analysis. *Information Processing in Medical Imaging. Springer International Publishing*, 2015. 2
- [38] D. Sun, T. G. van Erp, P. M. Thompson, C. E. Bearden, M. Daley, L. Kushan, M. E. Hardt, K. H. Nuechterlein, A. W. Toga, and T. D. Cannon. Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: classification analysis using probabilistic brain atlas and machine learning algorithms. *Biological psychiatry*, 66(11):1055–1060, 2009. 1
- [39] T. ur Rehman, E. Haber, G. Pryor, J. Melonakos, and A. Tannenbaum. 3d nonrigid registration via optimal mass transport on the gpu. *Medical image analysis*, 13(6):931–940, 2009. 2, 6
- [40] P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack. Alzheimer’s disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage*, 39(3):1186–1197, Feb 2008. 1
- [41] C. Villani. Topics in optimal transportation. *American Mathematical Society*, 58, 2003. 3
- [42] W.-Y. Wang, J.-T. Yu, Y. Liu, R.-H. Yin, H.-F. Wang, J. Wang, L. Tan, J. Radua, and L. Tan. Voxel-based meta-analysis of grey matter changes in alzheimers disease. *Translational neurodegeneration*, 4(1):1, 2015. 1
- [43] Y. Wang, X. Gu, T. F. Chan, P. M. Thompson, and S.-T. Yau. Volumetric harmonic brain mapping. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pages 1275–1278, 2004. 3
- [44] Y. Wang, J. Zhang, B. Gutman, T. F. Chan, J. T. Becker, H. J. Aizenstein, O. L. Lopez, R. J. Tamburo, A. W. Toga, and P. M. Thompson. Multivariate tensor-based morphometry on surfaces: application to mapping ventricular abnormalities in HIV/AIDS. *Neuroimage*, 49(3):2141–2157, Feb 2010. 1
- [45] F. Woermann, S. Free, M. Koepp, S. Sisodiya, and J. Duncan. Abnormal cerebral structure in juvenile myoclonic epilepsy demonstrated with voxel-based analysis of mri. *Brain*, 122(11):2101–2108, 1999. 1
- [46] I. C. Wright, P. K. McGuire, J. B. Poline, J. M. Traverre, R. M. Murray, C. D. Frith, R. S. Frackowiak, and K. J. Friston. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *Neuroimage*, 2(4):244–252, Dec 1995. 1